

PATENT ABSTRACTS OF JAPAN

W 1528

(11)Publication number : 2002-297322

(43)Date of publication of application : 11.10.2002

(51)Int.Cl.

G06F 3/06

G11B 20/10

G11B 20/12

(21)Application number : 2001-098499

(71)Applicant : HITACHI LTD
HITACHI SOFTWARE ENG CO LTD

(22)Date of filing : 30.03.2001

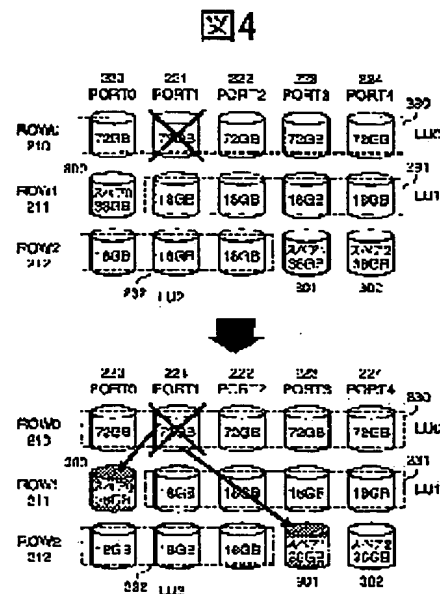
(72)Inventor : OKUMURA TOMOHIRO
TAKAMOTO KENICHI
MIZUMORI MOTOHIRO

(54) REDUNDANT DISK STORAGE DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To effectively use a spare disk drive without exchanging or scrapping by restoring data without regard to the capacity of a disk drive constituting RAID and that of the spare disk drive in a disk storage device having a RAID function and restoring data by the spare disk drive when a fault occurs in the disk drive.

SOLUTION: By referring to a whole disk drive resource information table and a spare disk drive resource information table, when a fault occurs in the disk drive constituting RAID, data is divided and restored by a plurality of spare disk drives of a capacity smaller than that of the disk drive where the fault occurs. When the fault of RAID occurs in data in the plurality of disk drives, the data is restored by a single spare disk drive.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the
examiner's decision of rejection or application
converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of
rejection][Date of requesting appeal against examiner's decision
of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-297322

(P2002-297322A)

(43) 公開日 平成14年10月11日 (2002. 10. 11)

(51) Int.Cl. ⁷	識別記号	F I	テマコード* (参考)
G 0 6 F 3/06	3 0 5 5 4 0	G 0 6 F 3/06	3 0 5 C 5 B 0 6 5 5 4 0 5 D 0 4 4
G 1 1 B 20/10 20/12		G 1 1 B 20/10 20/12	C

審査請求 未請求 請求項の数 5 O L (全 9 頁)

(21) 出願番号 特願2001-98499(P2001-98499)

(22) 出願日 平成13年3月30日 (2001. 3. 30)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(71) 出願人 000233055

日立ソフトウエアエンジニアリング株式会
社

神奈川県横浜市中区尾上町6丁目81番地

(72) 発明者 奥村 知弘

神奈川県小田原市国府津2880番地 株式会
社日立製作所ストレージシステム事業部内

(74) 代理人 100068504

弁理士 小川 勝男 (外2名)

最終頁に続く

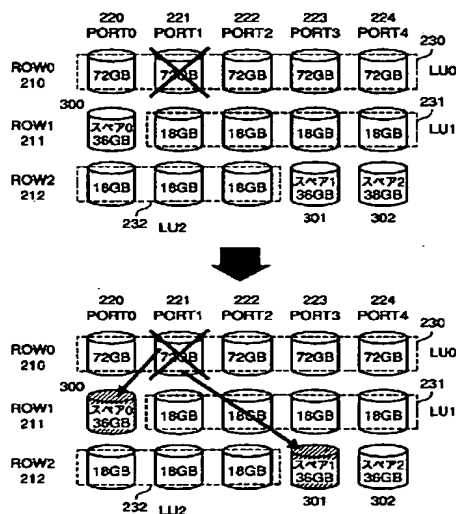
(54) 【発明の名称】 冗長性を有するディスク記憶装置

(57) 【要約】

【課題】 R A I D機能を有し、そのディスクドライブに障害がおこったときに、スペアディスクドライブに復元させるディスク記憶装置において、R A I Dを構成しているディスクドライブの容量、スペアディスクドライブの容量に関わらず、データを復元できるようにして、スペアディスクドライブを交換や破棄することなく有効に使用する。

【解決手段】 全ディスクドライブ資源情報テーブルとスペアディスクドライブ資源情報テーブルを参照して、R A I Dを構成しているディスクドライブに障害がおこったときに、その障害のおこったディスクドライブの容量よりも小さな容量の複数のスペアディスクドライブに、分割してデータを復元する。また、R A I Dの障害が複数のディスクドライブのデータにおこったときに、一台のスペアディスクドライブに復元する。

図4



【特許請求の範囲】

【請求項1】 ホストコンピュータと接続し、そのホストコンピュータの指示に基づいてディスクドライブの制御とデータアクセスと、各ディスクドライブをRAID (Redundant Array of Inexpensive Disks) として構築し制御する機能とを有し、スベアディスクドライブにデータを復元して動作させる冗長性を有するディスク記憶装置において、前記ホストコンピュータから各ディスクドライブにアクセスし、制御するための複数のポートとを有し、RAIDを構成しているディスクドライブに障害がおこったときに、その障害のおこったディスクドライブの容量よりも小さな容量の複数のスベアディスクドライブに、分割してデータを復元させることが可能なことを特徴とする冗長性を有するディスク記憶装置。

【請求項2】 ホストコンピュータと接続し、ホストコンピュータの指示に基づいてディスクドライブの制御とデータアクセスと、各ディスクドライブをRAID (Redundant Array of Inexpensive Disks) として構築し制御する機能とを有し、スベアディスクドライブにデータを復元して動作させる冗長性を有するディスク記憶装置において、前記ホストコンピュータから各ディスクドライブにアクセスし、制御するための複数のポートとを有し、RAIDを構成しているディスクドライブの複数の障害がおこったときに、それら複数のディスクドライブのデータを、一台のスベアディスクドライブに復元することが可能なことを特徴とする冗長性を有するディスク記憶装置。

【請求項3】 前記スベアディスクドライブの記憶領域を複数の分割して管理していることを特徴とする請求項2記載の冗長性を有するディスク記憶装置。

【請求項4】 データを復元後に、それらの複数のスベアディスクドライブを、一台のスベアディスクドライブとして、前記ホストコンピュータから通常アクセスすることが可能なことを特徴とする請求項1記載の冗長性を有するディスク記憶装置。

【請求項5】 データを復元後に、そのデータを復元した一台のスベアディスクドライブを、複数のスベアディスクドライブとして、前記ホストコンピュータから通常アクセスすることが可能なことを特徴とする請求項2および請求項3記載のいずれかの冗長性を有するディスク記憶装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、RAID機能を有し、スベアディスクドライブを設けて置くことが可能な冗長性を有するディスク記憶装置において、それらの複数のスベアディスクをデータディスクの容量に関係なく

有効利用できる可用性の向上させるのに用いて好適な冗長性を有するディスク記憶装置に関する。

【0002】

【従来の技術】 従来、ディスクアレイ装置など高度に信頼性を要するディスク記憶装置では、RAID技術に加えて、スベアディスクドライブを設ける技術が一般的に用いられている。このようなディスク記憶装置では、RAIDを構成しているディスクドライブの一台に障害がおこったときには、RAIDを構成している他のディスクドライブからデータを復旧して、RAIDの縮退状態での動作から、RAIDを構成しているディスクドライブが全て動作する通常アクセスの状態に復旧する。

【0003】 しかしながら、従来のディスク記憶装置では、スベアディスクドライブを用意するときには、RAIDが構成されているディスクドライブの容量と同容量か、または、それ以上の容量が必ず必要であった。RAIDが構成されているディスクドライブよりも、スベアディスクドライブの容量が小さい場合には、現在、設定されているスベアディスクドライブに、障害も無く使用可能なディスクドライブであっても使用することができないため、それ以上の容量のディスクドライブと交換しなければならなかった。このことは、顧客側にとって、使用できないディスクドライブが生じることになり、また、容量の大きなディスクドライブを用意しなければならないため高コストになると言う問題点があった。

【0004】

【発明が解決しようとする課題】 上記の様にRAID機能を有し、スベアディスクドライブを設けることのできるディスク記憶装置は、近年の情報化社会に伴い、要求された高性能、高信頼性、保守性の向上に応えるものであるが、可用性の向上も強く望まれている。

【0005】 上記来の技術で述べた通り、従来は、RAIDが構成されているディスクドライブ容量と同容量か、または、それ以上の容量がなければスベアディスクドライブとして設定できない方法が採られていた。

【0006】 しかし、この方法では、ディスクドライブよりも少ない容量のスベアディスクドライブが装置内にあった場合には、容量条件を満たした新たなディスクドライブと交換するということとなると、元のスベアディスクドライブが使用できるにも関わらず、取り外さなければならない、スベアディスクドライブ用の容量の大きな新たなディスクドライブを購入しなければならない、高コストになると言う問題点があった。

【0007】 本発明は、上記問題点を解決するためになされたもので、その目的は、RAID機能を有し、そのディスクドライブに障害がおこったときに、スベアディスクドライブに復元させるディスク記憶装置において、装置内の複数設けるスベアディスクドライブが、障害の

おこったディスクドライブよりも容量が小さいものであっても、そのスペアディスクドライブにデータを復元したり、一台のスペアディスクドライブに対して、複数のディスクドライブからのデータを復元できるようにして、スペアディスクドライブを交換や破棄することなく有効に使用させることを目的とする。

【0008】

【課題を解決するための手段】上記目的を実現するために、装置内の全ディスクドライブの容量や使用状態を把握するための全ディスクドライブ資源情報テーブルと、装置内に設定されている全スペアディスクドライブの容量や使用状態、回復元データディスクドライブ番号等を把握するためのスペアディスクドライブ資源情報テーブルを設ける。

【0009】そして、装置内の全ディスクドライブの情報は、全ディスクドライブ資源情報テーブルに反映される。また、スペアディスクドライブ資源情報テーブルには、スペアディスクドライブの状態が管理される。また、このスペアディスクドライブ資源情報テーブルにより、スペアディスクドライブを複数の領域に分けて、それにデータを復元することが可能になる。

【0010】この二つの情報テーブルを参照することにより、RAIDを構成しているディスクドライブに障害がおこり、スペアディスクドライブにデータを復元するときに、スペアディスクドライブが障害がおこったディスクドライブよりも、容量が小さいときであっても、複数のスペアディスクドライブにデータを復元させたり、一台のスペアディスクドライブに対して、複数の障害がおこったディスクドライブのデータを復元させることができ、RAIDを構成しているディスクドライブやスペアディスクドライブの容量に関わらず、ディスクドライブの有効利用が図れることになる。

【0011】

【発明の実施の形態】以下、本発明に係る一実施形態を、図1ないし図9を用いて説明する。

【0012】先ず、図1を用いて本発明のディスク記憶装置が使用されるシステムとディスク記憶装置の構成について説明する。図1は、本発明のディスク記憶装置が使用されるシステムとディスク記憶装置の構成について説明したブロック図である。

【0013】ディスク記憶装置110は、ホストコンピュータなどの上位装置100と接続されており、データを格納するディスクドライブ200と、上位装置100とディスクドライブ200との間の制御部120、ディスク記憶装置の操作、設定をおこなう制御コンソール180と、制御コンソール180と接続するためのLANインターフェース170から構成されている。

【0014】また、制御部120には、装置全体の制御をおこなうディスクアレイ制御部140、上位装置とのデータ転送に用いられるキャッシュメモリ150、上位

装置100とのデータ転送を制御する上位側転送制御部130、ディスク記憶装置110内のデータ転送を制御する下位側転送制御部160から構成されている。

【0015】ここで、本実施形態を記述するにあたりディスクドライブ200の物理的な並びをアレイとして捉え、横方向をROW(210~213)、縦方向をPORT(220~225)とし、ROW(210~213)について、それぞれにRAID5(パリティ付きストライピング)方式により冗長化された場合の論理ボリューム構成をLU(論理ユニット)(230~233)として管理するものとする。なお、本実施形態では、RAID5を前提とするが、他のRAID方式、例えば、RAID1(ミラーリング)などでも、本発明は有効である。

【0016】また、ディスクドライブ200の幾つかは、代替として使用するためのスペアディスクドライブ(300~304)として設定されている。

【0017】この設定は、制御コンソール180からグラフィカルなユーザインタフェースにより、おこなうことができる。

【0018】次に、図2および図3を用いて本発明のディスク記憶装置が参照するデータ構造について説明する。図2は、全ディスクドライブ資源情報テーブル500の構造を示す模式図である。図3は、スペアディスクドライブ資源情報テーブル600の構造を示す模式図である。

【0019】全ディスクドライブ資源情報テーブル500は、ディスクアレイ制御部140に存在し、装置内の全ディスクドライブの使用状態や容量を管理するテーブルであり、全ディスクドライブ単位管理領域510からなる。全ディスクドライブ単位管理領域510、図2に示すようにアレイとしてのデータ構造を持ち、装置内の全ディスクドライブがどのような状態か、そのディスクドライブの容量、またスペアディスクドライブへのデータ回復の進行状況を示すデータ回復済み容量を管理するところである。

【0020】使用状態510aには、未実装、実装(データディスク使用・スペアディスク使用)、未使用の四つの値が入り得る。未実装とは、最初から装置内に実装されていない状態、または障害が発生しディスクドライブを交換する際に装置から抜いた後の状態がこの未実装状態に該当する。未使用とは、ディスクドライブとしては、実装しているものの設定をしていないために、RAIDのディスクドライブとしても、スペアディスクドライブとしても使用されていない状態である。

【0021】全容量510bは、どのような容量のディスクドライブが実装されているかを管理する領域であり、スペアディスクドライブへのデータ回復の際の回復情報算出として使用する領域である。

【0022】回復済み容量510cは、このディスク

ライブに障害がおこったときに、全容量のうちで、現在どれだけ、スベアディスクドライブへのデータの回復をおこなったかを示す領域であり、ディスクドライブに障害がおこったときのスベアディスクドライブへのデータ復元時に用いる。

【0023】次に、スベアディスクドライブ資源情報テーブル600は、ディスクアレイ制御部140に存在し、スベアディスクドライブ(300~304)の状態を管理するテーブルであり、図3に示されているようにスベアディスクドライブ情報[0]~[n](610~670)からなる。

【0024】スベアディスクドライブ情報[0]~[n](610~670)には、スベアディスクドライブ(300~304)の状態の詳細情報が格納されている領域であり、使用状態611、スベア実ROW#612、スベア実PORT#613、スベア空容量614、空領域情報[0]~[n](615~617)、使用領域情報[0]~[n](618~620)からなる。

【0025】使用状態611は、そのスベアディスクドライブの使用状態を示し、未使用、一部使用、全使用の3つの値を持ち、データ回復の際のスベアディスクドライブ使用可否判断をおこなう領域である。

【0026】スベア実ROW#612と、スベア実PORT#613は、スベアディスクドライブの実装位置を示す領域であり、その値がこのスベアディスクドライブのアレイ上の位置を示している。

【0027】空領域情報[0]~[n](615~617)は、そのスベアディスクドライブがデータ回復に使用されていない領域を管理する領域であり、この領域が使用中であるかを示す使用中フラグ621、空領域の開始アドレス622、空領域の終了アドレス623からなる。使用中フラグ621には、空領域のときには、0が入るものとする。

【0028】使用領域情報[0]~[n](618~620)、そのスベアディスクドライブが、どれだけの領域をどこのディスクドライブのデータ復元用に使用されているかを管理する領域であり、その使用領域が使用中であるかを示す使用中フラグ624、使用領域の開始アドレス625、使用領域の終了アドレス626、どこのディスクドライブのデータ復元用に用いられているを示す、回復元ROW#627、回復元PORT#628からなる。

【0029】次に、図4ないし図6を用いてRAIDを構成するディスクドライブに障害がおこったときに、スベアディスクドライブにデータを回復するときの様子について説明する。図4ないし図6は、RAIDを構成するディスクドライブに障害がおこったときに、スベアディスクドライブにデータを回復するときの様子を示す模式図である。

【0030】この例では、図1に示されているディスク

ドライブのアレイの中で、図4のように(ROW0 210~ROW2 212)からなる3ROW構成で、ROW0 210に、LU0 230が、ROW1 211に、LU1 231が、ROW2 212に、LU2 232が設定されている。

【0031】最初に、図4に示すようにLU0 230のPORT1 221のディスクドライブ(72GB)に障害が発生したとする。この場合に、LU0は、RAID5で稼働しているので、その内の一台が稼働しなくても、パリティビットを利用して縮退状態で動作しているはずである。したがって、障害のおこっていないLU1の他の四台のディスクドライブからPORT1 221のディスクドライブのデータの内36GB分を、スベアディスクドライブ300に回復させる。このときには、ROW0 210 PORT1 221の回復済み容量510cは、36Gになっている。このデータの回復が終了後、引き続き、PORT1 221のディスクドライブからPORT2 221の回復していないディスクドライブの36GB分のデータをスベアディスクドライブ301に回復させる。このときには、ROW0 210 PORT1 221の回復済み容量510cは、72Gになる。

【0032】このような状態になれば、上位装置100からRAIDが回復した状態、すなわち、通常アクセスの状態としてアクセス可能になる。

【0033】このように障害がおこったディスクドライブの容量よりも、スベアディスクドライブの容量が小さいときであっても、複数のスベアディスクドライブに分割してデータを復元し、RAIDの復旧がおこなえることになる。

【0034】次に、図5に示すようにLU1 231のROW1 211、PORT4 224のディスクドライブ(18GB)に障害が発生したものとする。このときに、LU1の他の三台のディスクドライブから、PORT4 224のディスクドライブの全データを、スベアディスクドライブ302の前半分の18GBに回復させる。

【0035】次に、このような状態で、引き続き図6に示すようにLU2 232のROW2 212、PORT2 222のディスクドライブ(18GB)に障害が発生したものとする。このときに、LU2の他の二台ディスクドライブから、PORT2 222のディスクドライブの全データを、スベアディスクドライブ302の後半分の18GBに回復させる。

【0036】このような状態になれば、LU1とLU2の二つの論理ユニットが回復した状態になり、上位装置100から通常アクセスの状態としてアクセス可能になる。

【0037】このように、一台のスベアディスクドライブに対して、二台のディスクドライブのデータ復元し

て、複数の論理ユニットに対して、RAIDの復旧がおこなえることになる。

【0038】次に、図7および図8を用いてユーザが操作コンソール180から本発明のディスク記憶装置を操作するときのユーザインターフェースについて説明する。図7は、ディスクドライブの状態を画面に表示している操作コンソール180を示す模式図である。図8は、図4のディスクドライブの復元状態を画面に表示している操作コンソール180を示す模式図である。

【0039】図7に示されるように操作コンソール180の画面181には、ディスクドライブ200が、グラフィカルな状態で表示される。スベアディスクドライブを設定するときには、ターゲットなるディスクドライブをマウス183でクリックするなどして、設定できる。また、RAIDを構成しているディスクドライブ、スベアディスクドライブ、未使用のディスクドライブは、色分けして示せば分かりやすいインタフェースになる。

【0040】また、図8に示されるようにデータを復元状態のときにも、分かりやすいユーザインターフェースを提供することができる。

【0041】このとき、現在データの復元中であることを、矢印がアニメーションとして点滅する、または、矢印の長さが変化するなどの手法により示すことができる。

【0042】次に、図9を用いて本発明に係るディスク記憶装置のスベアディスクドライブへのデータ回復の手順について説明する。図9は、本発明に係るディスク記憶装置のスベアディスクドライブへのデータ回復の手順を示すフローチャートである。

【0043】まず、ディスクドライブの障害が発生し、スベアディスクドライブへのデータ回復を開始する（ステップ700）。最初に、ディスクアレイ制御部140は、どのスベアディスクドライブの、どの領域にデータ回復させればよいかをスベアディスクドライブ資源情報テーブル600を検索して決定する。次に、テーブル内の使用状態611、スベア空容量614、空領域情報[0]～[n]（615～617）の使用フラグ621、開始アドレス622、終了アドレス623、使用領域情報[0]～[n]（618～620）の使用フラグ624、開始アドレス625、終了アドレス626、仮想ROW#627、仮想PORT#628の情報を更新する（ステップ710）。そして、スベアディスクドライブ資源情報テーブル600に設定された情報を元にデータ回復を開始する（ステップ720）。

【0044】次に、使用領域（618～620のいずれか）内のデータ回復が完了したかを開始アドレス625と終了アドレス626から判断する（ステップ730）。

【0045】ステップ730の判断で、いまだ使用領域（618～620）分のデータ回復が終了していない場

合は、引き続きデータ回復を継続する（ステップ740）。

【0046】使用領域（618～620のいずれか）分のデータの回復が終了していた場合には、全ディスクドライブ資源情報テーブル500の全ディスクドライブ単位管理領域510にある回復済み容量を更新する（ステップ750）。

【0047】次に、データ回復対象となっているデータディスクドライブの全容量分のデータ回復が終了したかどうかを、全ディスクドライブ資源情報テーブル500の全ディスクドライブ単位管理領域510にある容量510bとステップ750で更新した回復済み容量510cとを比較して判断する（ステップ760）。

【0048】ステップ760の判断で、いまだ全データ回復が完了していなければ、次の使用領域（618～620のいずれか）のデータ回復を開始する。このように、ステップ730からステップ770までを繰り返すことによって、障害のおこったディスクドライブのデータの回復が完了する。

【0049】

【発明の効果】本発明によればRAID機能を有し、そのディスクドライブに障害がおこったときに、スベアディスクドライブに復元させるディスク記憶装置において、装置内の複数設けるスベアディスクドライブが、障害のおこったディスクドライブよりも容量が小さいものであっても、そのスベアディスクドライブにデータを復元したり、一台のスベアディスクドライブに対して、複数のディスクドライブからのデータを復元できるようにして、スベアディスクドライブを交換や破棄することなく有効に使用することができる。

【図面の簡単な説明】

【図1】本発明のディスク記憶装置が使用されるシステムとディスク記憶装置の構成について説明したブロック図である。

【図2】全ディスクドライブ資源情報テーブル500の構造を示す模式図である。

【図3】スベアディスクドライブ資源情報テーブル600の構造を示す模式図である。

【図4】RAIDを構成するディスクドライブに障害がおこったときに、スベアディスクドライブにデータを回復するときの様子を示す模式図である（その一）。

【図5】RAIDを構成するディスクドライブに障害がおこったときに、スベアディスクドライブにデータを回復するときの様子を示す模式図である（その二）。

【図6】RAIDを構成するディスクドライブに障害がおこったときに、スベアディスクドライブにデータを回復するときの様子を示す模式図である（その三）。

【図7】ディスクドライブの状態を画面に表示している操作コンソール180を示す模式図である。

【図8】図4のディスクドライブの復元状態を画面に表

示している操作コンソール180を示す模式図である。

【図9】本発明に係るディスク記憶装置のスペアディスクドライブへのデータ回復の手順を示すフローチャートである。

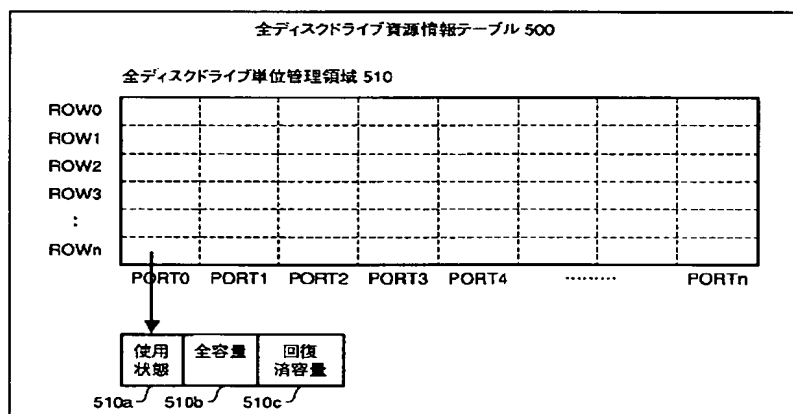
【符号の説明】

100…上位装置、110…記憶装置、120…制御部、130…上位側転送制御部、140…ディスクアレイ制御部、150…キャッシュメモリ、160…下位側転送制御部、170…LANインタフェース、180…制御コンソール、181…画面、182…キーボード、183…マウス、200…ディスクドライブ、210…ROW0、211…ROW1、212…ROW2、213…ROWn、220…PORT0、221…PORT1、222…PORT2、223…PORT3、214…PORT4、215…PORTn、230…LU0、231…LU1、232…LU3、233…LU_n、300…スペアディスクドライブ0、301…スペアディスクドライブ1、302…スペアディスクドライブ2、

303…スペアディスクドライブ3、304…スペアディスクドライブ4、500…全ディスクドライブ資源情報テーブル、510…全ディスクドライブ単位管理領域、600…スペアディスクドライブ資源情報テーブル、610…スペアディスクドライブ情報[0]、611…使用状態、612…スペア実ROW#, 613…スペア実PORT#, 614…スペア空容量、615…空領域情報[0]、616…空領域情報[1]、617…空領域情報[n]、618…使用領域情報[0]、619…使用領域情報[1]、620…使用領域情報[n]、621…使用中フラグ、622…空開始アドレス、623…空終了アドレス、625…使用開始アドレス、626…使用終了アドレス、627…回復元仮想ROW#、628…仮想PORT#、630…スペアディスクドライブ情報[1]、640…スペアディスクドライブ情報[2]、650…スペアディスクドライブ情報[3]、660…スペアディスクドライブ情報[4]。

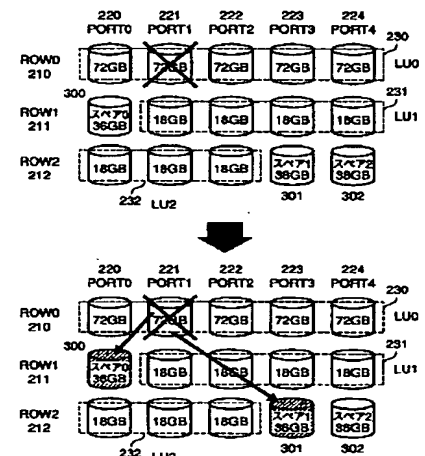
【図2】

図2



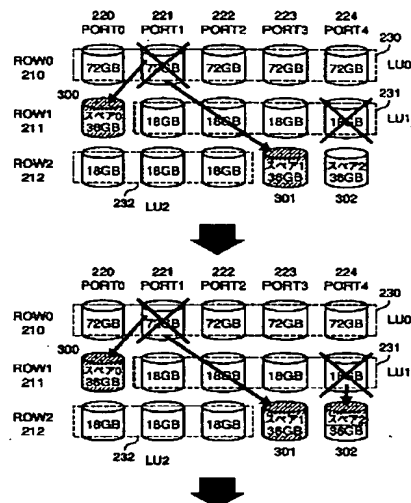
【図4】

図4

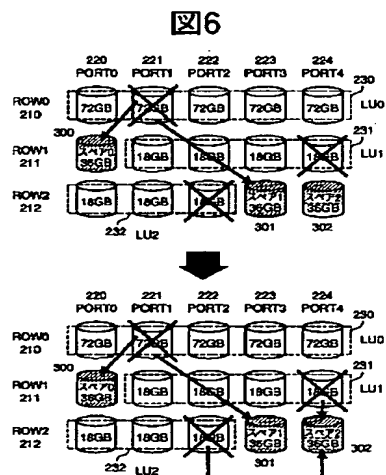


【図 5】

图5

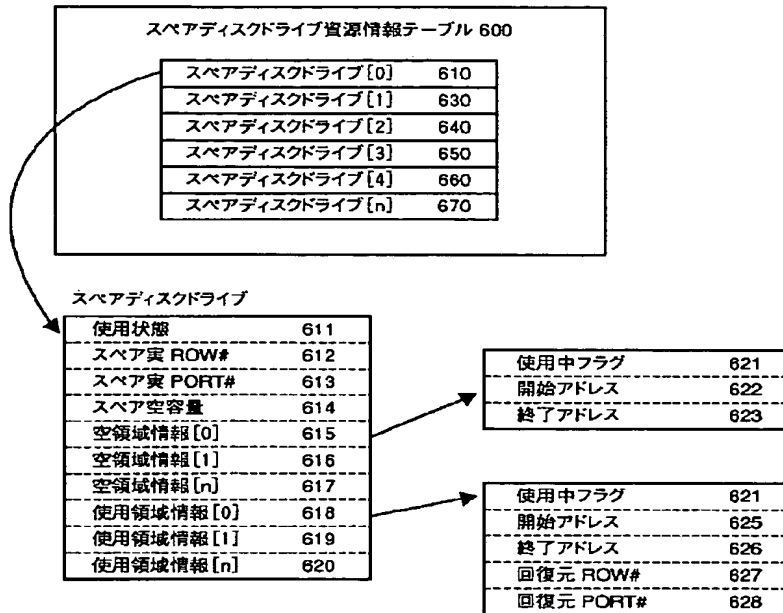


【図 6】



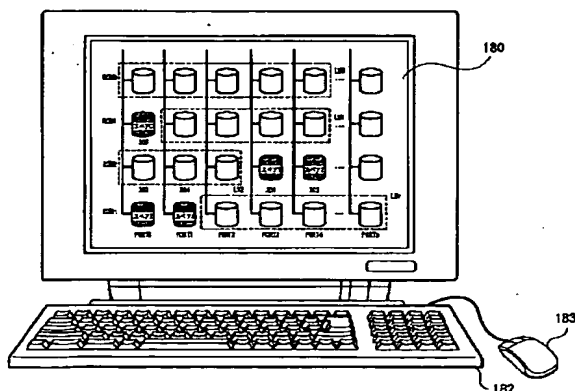
【図3】

図3



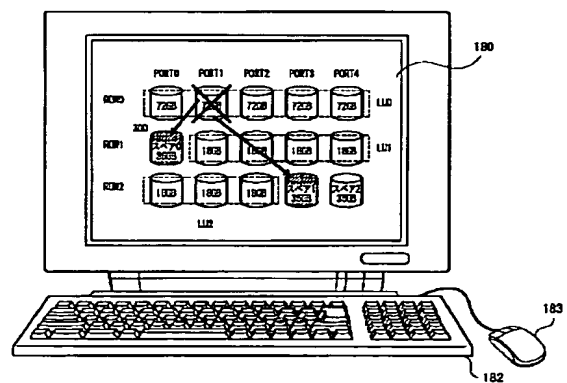
【図7】

図7



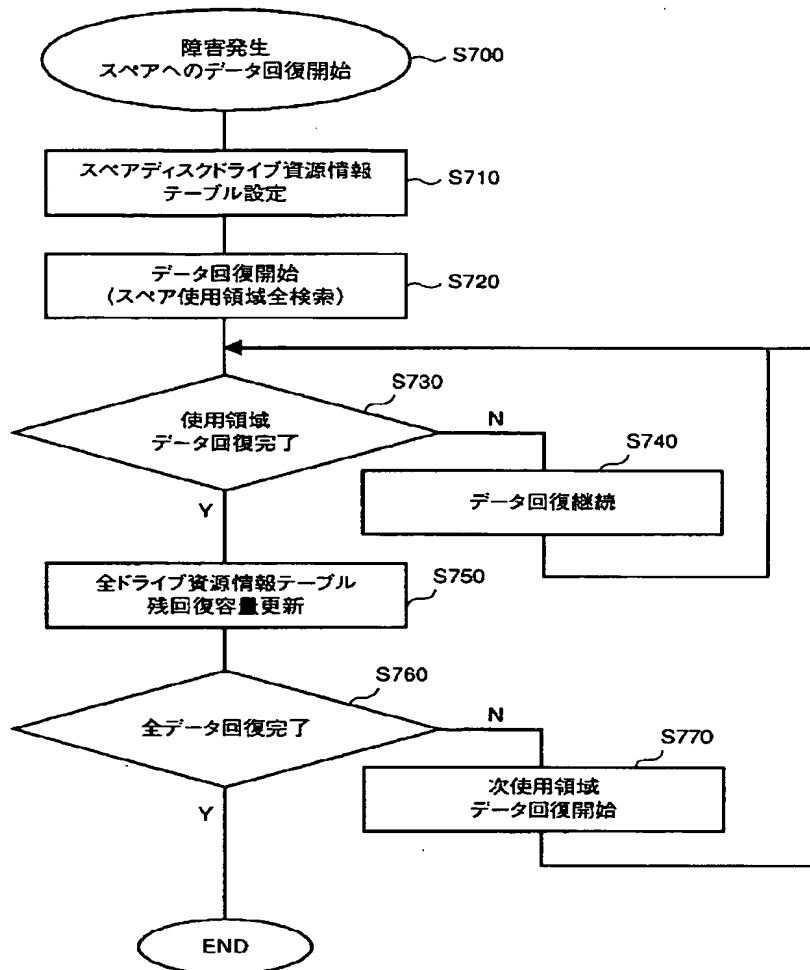
【図8】

図8



【図9】

図9



フロントページの続き

(72)発明者 ▲高▼本 賢一

神奈川県小田原市国府津2880番地 株式会
社日立製作所ストレージシステム事業部内

(72)発明者 水森 源宏

神奈川県横浜市中区尾上町6丁目81番地 日
立ソフトウェアエンジニアリング株式会
社内

45

Fターム(参考) 5B065 BA01 CA12 CC08 EA02 EA18

EA24

5D044 BC01 CC04 DE62 GK11 GK19